Fundamentals of Multimedia Computing

Chapter 5: More Sensors

Authors:
Gerald Friedland
and
Ramesh Jain

Draft for Comments

More Sensors

The sensors for sound and light described in the previous two chapters have by far been the most important for multimedia computing in the last decades. The reason for this is that both audio and video are best for conveying information in a communicative sense for the tasks typically performed with a computer, i.e. sound is the human-preferred way of communication and light serve illustrative purposes supplementing the need for language-based description of a state of the world. New or different tasks might use different sensors, however. For example, in dating, communication probably occurs on many other levels, such as scent, touch, and taste. In this chapter we will therefore initially discuss other sensors in a general way, including sensors that are not part of the human sensoric system and then touch on another important topic: Multimodal integration, i.e. the integration of different sensors into one integrated sensation.

Properties and Types of Sensors

In general, a sensor is a device that measures a physical quantity and converts it into a signal which can be read by an observer or by an instrument. Whether the sensor is human-made or biological nature does not matter. An ideal sensor has the following properties: It is sensitive to the measured property, insensitive to any other property, and does not influence the measured property. Of course, no perfect sensor exists, since the laws of physics state that energy is conserved and any sensors needs a transfer of energy to function, after all the photos absorbed in the retina of a human eye interfere with the universe in the sense that the photos would not be absorbed at all or would be absorbed by a different object. In practice, however, these rules help optimizing the development of sensors. It also helps to understand common deviations from ideal sensors, in fact, it's these deviations that multimedia computing utilizes for compression, corrects when reproducing signals, and analyses for content analysis. Many of them have already been discussed in previous chapters:

If the sensor is not ideal, several types of deviations can typically be observed: No sensor has unlimited range. Therefore the sensor might saturate, i.e. while the ideal measurement response would suggest a further increase in output, the sensor outputs a maximum saturated value and/or breaks (e.g compare human ears exposed to too loud noises). the lower bound of the range is defined by the minimum amount of input that can be clearly distinguished from no input. If the output is not zero when the input is zero, the sensor is called to have an offset or bias. The sensitivity may in practice differ from the measurement function specified. Ideally a sensor should respond linearly to the measured entity, i.e. the sensitivity should be constant. This error is therefore often described as "non-linear behavior". Often sensors are tuned to be behave linear inside an operational range. The error may be dynamic in that the sensor might behave differently based on time or other influencing factors that vary independently from the measured entity. A changing sensitivity given a constant signal is called drift. Most sensors have long-term drift due to aging. A random deviation from the measurement function is called noise. Hysteresis is a deviation over time: When the measured entity reverses direction (e.g. gets higher instead of lower) but the response of the sensor has a finite lag in response it may create a different offset in one direction than in the other. Figure 1 illustrates the concept. Errors created due to sampling are called digitization error and aliasing errors and are discussed in depth in Chapter XXX. The

resolution of a sensor is the smallest change it can detect in the quantity that it is measuring. Resolution might also behave non-linearly.

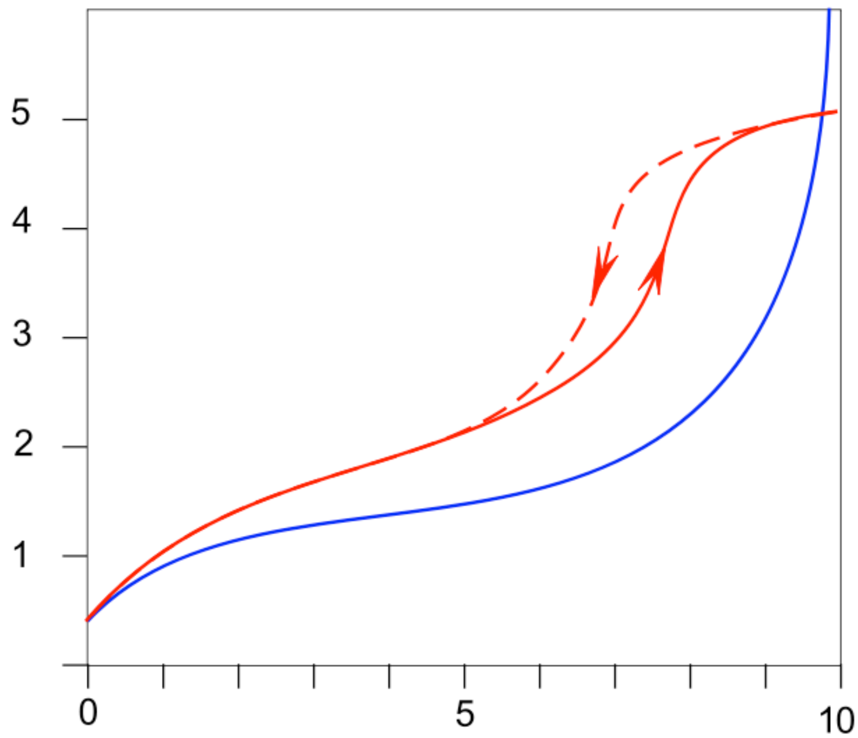

Figure 1. A general example for the concept of hysteresis: The input for the sensor is shown in the bottom curve (blue). The output on the upper curve. The curve behaves differently when the measured entity decreases compared to when it increases (arrows).

Despite the difficulties in creating ideal sensors, different types of sensors have evolved in nature of which some of them can be (to some extend) reproduced artificially. We already discussed sensors for light and sound. In addition, many animals have sensors for temperature, magnetic and electrical fields, gravity, humidity, vibration, pressure, certain chemicals (smell) and other properties of their environment. Many nerve systems also sense internal aspects of the body, such as motion, stretch, positioning of body parts (proprioception), and unexpected states (surprise). Many chemical sensors are also in place, for example for measuring the amount of Carbondioxide in our lungs. In addition to these, many more artificial sensors have been produced to sense physical as well as chemical phenomenons, such as radiation, force, flow, seismic activity, and so on. In multimedia computing we mostly care about sensors important for human communication, therefore these sensors are not discussed in this book.

Also, while many researchers believe that human communication involves all sensors, including smell, touch, and taste, only sound and light sensors could be developed well enough be in every day's practical use. Please refer to the references about the artificial nose and electrical tongue projects. One sensor that is in increased use in multimedia computing, however, is touch.

Haptic technologies target tactile sensors both in humans (as emitter of the sensations) as well as for the computer (as sensor). Currently, the predominant use of haptics is forced-feedback in computer game controllers, servo mechanisms for aircrafts, and medicine. These devices use electro motors to induce a vibration or another physical motion to communicate with the tactile sensors in the human body (mostly in the hands and arms). The device is called actuator. On the sensor side, pressure sensitive material can be used, such as piezo crystals. These crystals emit electrical power proportional to the force that is used to squeeze them. Other sensors, such as haptic gloves, allow normal usage of the hand but capture the muscular state of the fingers and other parts of the hand. Often, a remote robotic hand can then be operated by reproducing the state of the operators hand on the remote end. A major issue with haptic sensors is to tune the feedback and or sensor so that the user experience is intuitive and natural to avoid having to learn how to use the systems. While this is easily achieved with conventional video and audio, haptic sensors often require the combination of different sensors and feedback -- which leads us to the topic of multimodal integration.

Multimodal Integration

Multimodal Integration, also known as multi-sensory integration, describes the phenomenon that that the human brain is able to integrate different sensory modalities, such as sight, sound, and touch, into a perceptual experience that is coherent and unified. Experiments show that by considering input from multiple sensors, perceptual problems can be solved more robustly and even faster (see references). The phenomenon is not yet completely understood, but yet somewhat fundamental to the success of multimedia systems. Moreover, as we see in Chapter XXX, in computer science, synergistic use of data encoded for different human sensors has not yet lived up to its promise for most cases. Multimedia computing strives to imitate the properties of multimodal integration regardless of the incomplete understanding of the mechanisms in the brain, e.g. multimedia content analysis (as described in Chapter XXX) tries to use audio and video in combination to gain accuracy, robustness, and sometimes speed.

Multimodal integration is still an active area of research and the mechanisms behind it are not well-understood. In the following, we will describe some well-known observable phenomena that may not help in the understanding of the phenomenon but also in the design considerations for multimedia systems.

One of the most fundamental phenomenons is the McGurk effect. Experiments have indicated that two converging sensory stimuli can produce a perception that is not only different in magnitude than the sum of the two individual stimuli but also quite different in quality. The classic study, that is now known as McGurk effect, dubbed a person’s acoustic phoneme production with a video of that person speaking a different phoneme. The result was neither the perception of the visual or acoustic pronunciation but the hearing of a third, different phoneme. McGurk and MacDonald explained in their article in 1976 (see references) that phonemes such as ba, da, ka, ta, ga, and pa can be divided into two groups: Phonemes that can be visually confused, i.e. (da, ga, ka, ta), and phonemes that can be acoustically confused (ba and pa). As a result, the combination of the visual and acoustically confused phonemes results in the perception of a different phoneme, e.g. when ba is uttered and dubbed on a video that shows the uttering of ga, which is processed visually through lip reading, the visual modality sees ga or da, and the auditory modality hears ba or da, combining to form the perception of da.

Another important effect is known as ventriloquism. It describes the situation in which acoustic tracking of the origin of a sound is shifted towards the visual modality. In conditions in which the visual cue is unambiguous the perception of the visual location overrides the acoustic location. This effect is broadly used by artists throughout the world. Ventriloquists manipulate their own voice so that it appears that the voice is coming from elsewhere, usually a puppet.

An almost "magic" effect is called Body Transfer Illusion. The original, so-called rubber hand experiment was performed by Botvinick and Cohen in 1998 (see references). Human participants were sitting in front of a screen showing a dummy hand being stroked with a brush while they were feeling a series of synchronized and identical brushstrokes applied to their own hands, hidden from their own views. The result was that if the visual appearance and the dummy hand is similar to the participants hand in appearance, position, and location, the human subject may feel that the touches on their own hand are coming from the dummy hand. Furthermore, several participants reported that they felt the dummy hand to be their own hand. Virtual reality applications take advantage of this effect and try to induce the perception of owning and controlling somebody else's body (usually an avatar) by applying visual, haptic, and optionally proprioceptual stimulation synchronously.

The brain takes advantage of multimodal integration in different ways. Probably, the two most important are the decrease of sensory uncertainty and the decrease of reaction time. Experiments have shown that uncertainty in sensory domains results in an increased dependence of multisensory integration. If a person sees something moving in a tree and isn't sure whether it is a bird or a squirrel, the natural reaction is to listen. If a chirp sound is emitted and localized by the brain to be coming from said tree then it is taken as a proof for the creature being a bird. Hence, the lack of visual information is augmented by acoustic information. The Hershenson experiments (see references) also showed that responses to multiple simultaneous sensory stimuli can be performed faster than responses to the same stimuli presented in isolation. Participants were presented a light and tone simultaneously and separately, and were asked to respond as rapidly as possible to them by pressing a button. Reaction time differed with varying levels of synchrony between the tone and the light. The former result is, however, hard to generalize as multiple synchronous stimuli might also cause the opposite effect -- as discussed in the next Section.

Split Attention

The opposite effect of multimodal integration is called split attention. It is apparent when the same media (e.g. visual and visual) is used for different types of information at the same time. To pay attention to the content, one must split the attention between these materials to understand and use the materials provided. Split attention is not to be confused with distraction. Distraction is usually described as being caused by the lack of ability to pay attention to a particular subject due to lack of interest in the object or the great intensity, novelty or attractiveness of something other than the object of attention. The two problems are related, of course. However, split attention is a phenomenon that caused by the lack of integration of the object to be paid attention to. Of course, attention is still a topic of research and it is still an open question whether it can be split in order to attend to two or more different sources of information simultaneously. The topic has been discussed by psychologists and neuroscientists for decades. Most researchers, however,

now accept that attention can be split at the cost of cognitive overhead. It is this cognitive overhead that is to be avoided when designing multimedia systems. See the references for more information.

Figure 2 shows an example of multimedia system known to have caused a split attention problem. The E-Chalk lecture recording system showed dynamic board content in addition to a video of the lecturer. In a typical E-Chalk lecture there were two areas of the screen competing for the viewer's attention: the video window showing the instructor and the board or slides window. Several researchers (see references) tracked the eye movements of students while watching a lecture recording that contains slides and an instructor video in a similar setup as shown in Figure 2. Measurements showed that a students often spends about 70 percent of the time watching the instructor video and only about 20 percent of the time watching the slides. The remaining 10 percent of the eye focus was lost for activities unrelated to lecture content. When the lecture replay only consists of slides and audio, students spend about 60 percent of the time looking at the slides. Of course, there is no other spot to focus attention on in the lecture recording. The remaining 40 percent, however, were lost due to distraction. The example illustrates how two areas of a screen can be competing for attention, which causes cognitive overhead.

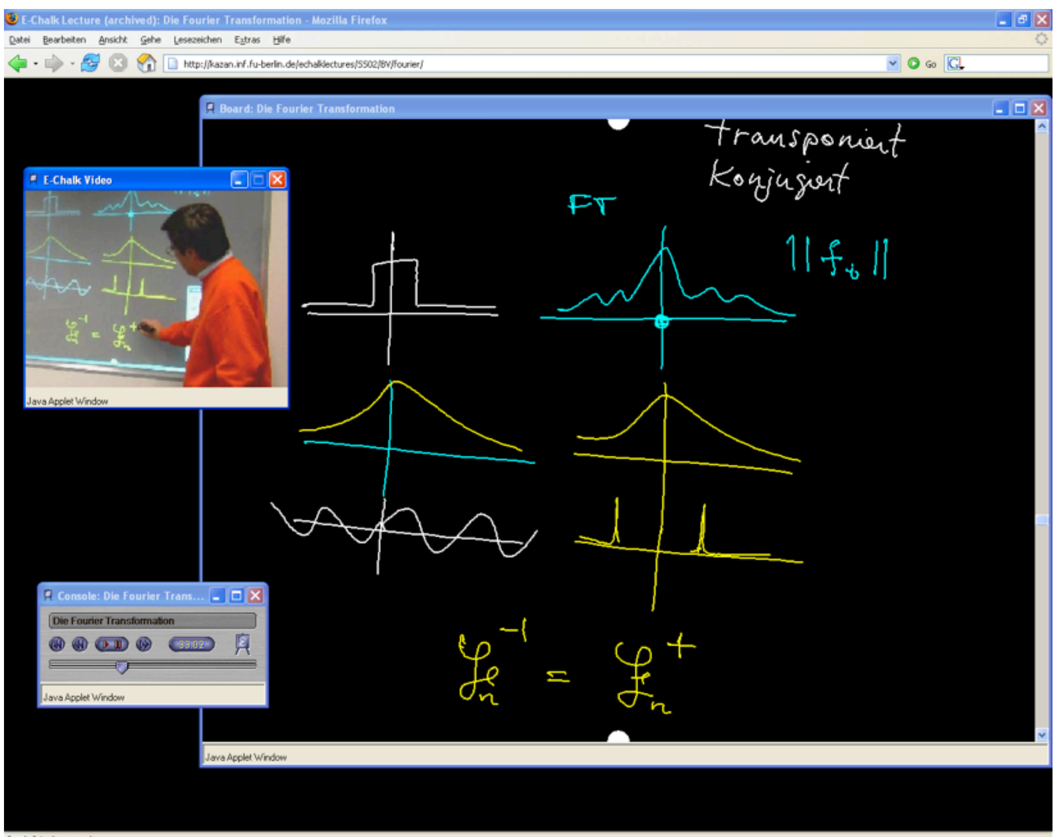

Figure 2. An example of a split-attention problem: The lecturer on the left is shown as a video, the dynamic board content he creates is shown on the right. The additional video window is used to convey gestures and finger pointing. However, presenting the lecturer in a second window

causes cognitive overhead usually referred to as split attention because the viewer switches between the content on the board and the lecturer on the left.

Here is another example: So far, we have described sensory phenomena on a high level and discussed important properties of the human system. From now on. This book will now begin to dig deeper into what it means to process sensory information in different ways using computers. While reading the next chapters you might find it useful to go back to these introductory chapters from time to time to remind yourself of the fundamentals described here in. You will do so at the cost of some cognitive overhead, a phenomenon called split attention...

Literature

- C. A. Grimes, E. C. Dickey, and M. V. Pishko (2006), Encyclopedia of Sensors (10-Volume Set), American Scientific Publishers.
- Gemma Calvert, Charles Spence, Barry E. Stein: The handbook of multisensory processes, MIT Press, 2004.
- Baars, B. J. (1988): A Cognitive Theory of Consciousness. Cam- bridge University Press, Cambridge, UK, 1988.

Research Articles

- A. Riul, Jr, R. R. Malmegrim, F. J. Fonseca, and L. H. C. Mattos: An artificial taste sensor based on conducting polymers, Biosensors and Bioelectronics, Volume 18, Issue 11, October 2003, Pages 1365-1369.
- Krishna Persaud, George Dodd: Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose, Nature 299, 352 - 355 (23 September 1982).
- M. Hershenson: Reaction time as a measure of intersensory facilitation. J Exp Psychol, 63:289–93, 1962.
- H. McGurk and J. MacDonald: Hearing lips and seeing voices. Nature, 264(5588):746–48, 1976.
- Hahn, S. and Kramer, A. F. (1998). Further evidence for the division of attention among non-contiguous locations. Visual Cognition, 5(1-2):217–256.
- Narcisse P. Bichot and Kyle R. Cave and Harold Pashler (1999). Visual selection mediated by location: Feature-based selection of non-contiguous locations. Perception & Psychophysics, 61 (3):403–423.
- Chandler, P. and Sweller., J. (1992). The Split Attention Effect as a Factor in the Design of Instruction. British Journal of Education Psychology, 62:233–246.
- Sweller, J., Chandler, P., Tierney, P., and Cooper, G. (1990). Cognitive Load as a Factor in the Structuring of Technical Material. Journal of Experimental Psychology: General, 119:176–192.
- Glowalla, U. (2004). Utility und Usability von E-Learning am Beispiel von Lecture-on-demand Anwendungen. Fortschritt-Berichte VDI, 22(16):603–621.

- Mertens, R., Friedland, G., and Krueger, M. (2006). To See or Not To See: Layout Constraints, the Split Attention Problem and their Implications for the Design of Web Lecture Interfaces. In Proceedings of the AACE E-Learn – World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, Honolulu, Hawai, USA.

- Botvinick, M. & Cohen, J. D. (1998). Rubber hand 'feels' what eyes see. Nature, 391, 756.

Exercises

1. Traditional wisdom gives humans five senses: vision, hearing, touch, smell, and taste. Explain why this number is wrong and discuss why it is not easy to define what a sense is.
2. Habituation describes an effect where repeated exposure to a stimulus leads to decreased responding. Provide examples of human sensors that show habituation.
3. Explain why haptic sensors in virtual reality require physical robustness beyond what is often technically achievable.
4. Explain why the McGurk effect is generally not a problem for the dubbing of TV shows and movies into a different language.
5. Find at least one more example for the split attention issue in daily life. Propose a solution.